PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Protocol article: Impact of bariatric surgery on neural food processing and cognition – an fMRI study
AUTHORS	Schulze, Marcel; Sörös, Peter; Vogel, Wolfgang; Münte, Thomas; Müller, Helge; Philipsen, Alexandra

VERSION 1 – REVIEW

REVIEWER	Baoci Shan Beijing Engineering Research Center of Radiographic Techniques and Equipment, Institute of High Energy Physics, Chinese Academy of Sciences, Beijing
REVIEW RETURNED	27-Mar-2018

GENERAL COMMENTS	In this protocol article, the authors want to investigate the relationship between the brain dysfunction, cognitive impairments, and hormones disruption in obese subjects, as well as the effect of RYGB to these indexes. It is helpful for us to understand the neural-mechanisms of obesity and its treatment. Before publication, some problems should be revised. 1, Please add some detail information to your Methods part. Are the obese subjects with diabetes be enrolled in your study? How could you ensure the control group be healthy, medical examination, medical history investigation, or any other route ? 2, In line 29 to 30, " the subjects will be required not to eat 3h prior to the measurement", why the time of fasting is 3 hours? Please give out the criterion. In "The energy intake earlier that day and the day before the measurement will be recorded", "that day" point to which day? 3, What's your fMRI sequence parameters? In general, the TR of fMRI for 3T is 2s. If this, each image should be presented for 1 TR and with an ISI 1 TR at least. 4, In the MRI data analysis, the timing correction should be before the motion correction. And the spatial normalization step should be after the head motion correction but before the spatial smooth. In line 15 to 18, what do you mean by saying " The temporal derivative of the hemodynamic response function will be included as regressor of no interest to model" ? I suggest to change it to "Some nuisance covariates such as head motion parameters, global mean signal, white matter signal, and cerebrospinal fluid signal were regress out to reduce the no interest signals." 5, In line 27 to 28, "For thresholding and correction for multiple comparisons, cluster-wise thresholding will be used" The thresholding is for multiple comparisons correciton. What's your thresholding? I suggest to change it to "For correction of multiple comparisons, clusters were threshold at (your value). And put this sentence back to the group comparisons. Also, please give out your P values.

	6, Please noted the order of the data analysis steps.

REVIEWER	Lyle Wiemerslage, PhD and Gaia Olivo, MD Uppsala University, Department of Neuroscience, Sweden
REVIEW RETURNED	13-Apr-2018

GENERAL COMMENTS	Firstly, I want to commend the authors as well as the journal for putting this protocol up for review prior to implementation! It is much more worthwhile to discuss the theory in the discussion than complain about what could/should have been done in the methods. Moreover, having a publication dedicated specifically to the planning/methods of a study will be highly useful to groups pursuing similar work and help them plan their own research accordingly. I cannot say enough good things about this process! So I'll just leave it at that. Now onto the review – while I'm completely confident in the authors' ability to do good science and I look forward to the future results, I do have a number of questions and what I hope the authors will consider as constructive criticisms. Also, I should mention that the review was performed together with a close colleague. We have both reviewed the protocol independently and then discussed. While our assessments were similar, and we are in complete agreement with our consolidated comments below: 1. The protocol was unclear as to if the patients will be fasted or sated during scanning? This is an important factor to consider as past studies have found differences between prandial states. In any case, hunger should be assessed prior to scanning and prandial state should be carefully controlled between all groups. Moreover, fasting should not be equated with hunger. Just because someone fasted overnight, does not mean they are hungry. Perhaps scans in the early afternoon after at least a 4-hour fast are a superior fasting/hungry condition compared to overnight fasts. We shouldn't necessarily measure hunger in the brain the same way we measure blood sugar. I encourage the authors to consider what exactly are the phenomena of fasting versus hunger, and when each is best displayed/measurable: time of day, time from previous meal, size of previous meal, etc. Likewise, if the authors intend to scan in a sated state, then they carefully consider the meal size, timing, palatability, etc. Some studies have used small, set meal sizes, while other studies have employed meals with caloric contents based on body size. I again encourage the authors to likewise think deeply about what "sated" means and how it is best displayed/measured for the purposes of their study. 2. Have the authors considered scanning in both fasting and sated states (both before and after surgery)? It is twice the scanning, but it exponentially increases the potential findings of the study. The resting-state fMRI would be the most important to consider for these extra scans. 3. Previous studies have found sometimes conflicting results between males and females in fMRI studies regarding food cues. And with only 25 obese participants planned, it seems prudent to consider enrolling only one sex. Admittedly, this is not ideal, but it simplifies the analysis and leaves less room for argument. I would consider females the more interesting/important choice of study (as males have been typically preferred in past studies). Regardless of what the authors decide, the female hormonal cycle should be controlled unless the authors can adequately explain that it is not

	expected to be an issue. Also, if the authors ignore this complaint, they should be careful to split the balance between males and females. 4. The inclusion/exclusion criteria are rather vague: a. Could the authors provide the full list of inclusion/exclusion criteria? b. Will medication status be considered a criterion? Psychopharmacological treatment has been mentioned as an exclusion criterion, but what about medications that might affect hormonal levels – a secondary outcome of this study? c. Participants will be assessed for comorbidities by administering the SKID-I. But how will the results of this assessment be used? In case of comorbidites, will the participant be included in the study all the same, or will he/she be rejected? d. Neurological disorders should be an exclusion criterion. 5. Haven't hypotheses #1 and 2 already been demonstrated in previous works? How exactly will this study build on these past findings? 6. Regarding hypotheses #3 and 4, I don't understand why memory performance is of concern (other than that it has been tested previously). Eating behaviour is the primary psychological element we are interested in. So shouldn't tests specific to eating behaviour be considered primary endpoints? Indeed, working memory performance and executive functions (as measured via the Trail Making Test; J Clin Exp Neuropsychol. 2000 Aug;22(4):518-28.) have both been found to be affected in obesity by meta-analyses (Cook, Rebecca et al. Obesity Research & Clinical Practice , Volume 8 , 21; Veronese et al., Neurosci Biobehav Rev. 2017 Jan;72:87-94. doi: 10.1016/j.neubiorev.2016.11.017.). 7. Hypothesis #6 is too vague. How exactly do the authors expect hormones to correlate with the other variables measured? Also, what exactly is of consequence in these correlations regarding behaviour or neural activity? How will these insights lead to improved treatments? 8. Why is the follow up 12 months? I suppose it controls for the time of year, but shouldn't results from the surgery be measurable before that? Some studies have seen differences in neural activity as early as one month after surgery (perhaps that's too soon after the recovery). But the goal is to simply measure patients after they have lost weight (or failed to lose weight), correct? Why needlessly make the study longer than it needs to be and potentially lose patients to follow up? Indeed, a year after surgery may be too long – perhaps some patients have an initial weight loss and then rebound. The literature should be consulted for an optimal follow-up time point. Perhaps weight can be tracked via a patient diary? 9. The authors should consider collecting blood samples for future genetic analysis, as there are several other studies in the literature that they could compare results. Variants for the FTO gene have been the most popular to date. 10. Regarding the images shown to the participants during the fMRI task, how exactly are the images chosen? Are the food images highly palatable, equal in caloric content, macronutrients, taste, perceived healthiness, etc.? Likewise, are the control images

	matched for features relative to the food images: colors, shapes, etc.? How exactly is it determined that a food is high-calorie versus low-calorie? Also, something can be high-calorie but still perceived as relatively healthy, which may not be the intended stimulus to the participant. The authors should consider what feelings/associations will be intentionally or unintentionally elicited to the participants. For example, an image of a birthday cake may remind someone that their birthday was forgotten and unintentionally illicit a neural response based on sad emotions. Perhaps a 1-second interval is too fast for emotive processing, but that should be adequately explained, if so. 11. The independent-component analysis for the resting-state fMRI is sound, but is also a somewhat curious choice, considering there are so many brain areas that qualify for a priori testing. Perhaps that is why the authors chose this method (too many brain areas)? Regardless, I would be interested to know their reasoning/comparisons to the other available methods for analysing resting-state data. 12. While the detailed description of the statistical analyses is appreciated, the authors should assess the validity of the assumptions before choosing the appropriate tests. This can only be done after data collection. For example, they mention that Student's t-test will be used for hormonal measurements. The t-test, as all parametric tests (ANOVA included) have several assumptions, mainly the normality of the residuals and homogeneity of variance between groups. 13. Regarding the neuroimaging protocol, are the scans assessed by a neuroradiologist to ensure the absence of structural anomalies? This is usually a standard procedure, but still important to say so – especially for a "methods" manuscript such as this. 14. Concerning the resting-state fMRI, the authors might consider having the participants fixating on a cross during the acquisition, as this protocol has been demonstrated to have a better test-retest reliability (Patriat, Neuroimage. 2013 Sep;78:463-73. doi: 10.1016/j.neuroimage.2013.04.013.). This could be useful, as this is a longitudinal study.

VERSION 1 – AUTHOR RESPONSE

Reviewer#1

Dear Mr. Baoci Shan

1. Please add some detail information to your Methods part. Are the obese subjects with diabetes be enrolled in your study? How could you ensure the control group be healthy, medical examination, medical history investigation, or any other route?

Reply: Comorbidities such as diabetes mellitus, hypertension, sleep apnea, and dyslipidemia will be no exclusion criteria to get enrolled in the study since these are obesity-related comorbidities. Suffering from one of the comorbidities (plus having a BMI>35kg/m2) is a prerequisite to get a bariatric surgery. This has been noted in the manuscript

Concerning your question of ensuring the healthiness of the control group, we intend to perform the same behavioral assessment in obese patients and healthy controls. This assessment includes questionnaires of psychiatric comorbidities (here the structured clinical interview will be administered), depression (assessed through Beck Depression Inventory), ADHD-symptoms (Conners' adult ADHD

self-rating scales), recording of obesity-related comorbidities, and known neurological disorders in the past (assessed through a demographic questionnaire).

To cope with the raised issues, we have therefore now done the following:
(1.) We explicitly wrote down that obesity-related comorbidities are no exclusion criteria.

2. In line 29 to 30, " the subjects will be required not to eat 3h prior to the measurement", why the time of fasting is 3 hours? Please give out the criterion. In "The energy intake earlier that day and the day before the measurement will be recorded", "that day" point to which day?

Reply: We thank the reviewer for raising this point. We discussed the fasted period again and realized that 3 hours might be too short; therefore, we decide to extend the fasted period to 4 hours. The time interval of being in a fasted state 4 hours prior to the measurement was aggregated from a meta-analysis on neuronal food processing1. Here, in table 1 a comparison of studies comparing food vs. nonfood stimuli is displayed. For instance in the study of Uher et al. 2006 the experimental group had a mean fasted period of 3.5 hours (also showing objects as control stimuli). Additionally, it is known that cognitive performance is modulated by the postprandial blood glucose profile, with interpersonal-variations in outcomes up to 225 min postprandial. To overcome these variations, we decided to have a fasted state of 240 min postprandial. We are, however, aware that interpersonal differences could still exist, especially in the presence of diabetes mellitus (which is no exclusion criterion). To control these potential confounds, we include a hormonal analysis of GLP-1 (as a marker of insulin secretion and therefore lowering the blood glucose level) and PYY2. Further, we plan to correlate the hormones with the cognitive performance and with the brain responses to food-stimuli.
To cope with the raised issues, we have therefore now done the following:
(1.) We state the measurements on the variables assessing hunger and energy intake more specifically. We therefore changed the sentence "The energy intake earlier that day and the day before the measurement will be recorded" to “The participants will be in a fasted state 4 hours prior to measurement. All the measurements will take place at around the same time of the day, which will be between 3pm, and 5pm. The current state of hunger at the recording time will be assessed on a ten-point-Likert scale. Additionally, the time from the previous meal and portion size of previous meal, will be recorded. To have a measurement on food consumption before entering the study, the participants will be asked to complete a food diary 1 week prior to the study in order to record what has been eaten, at which time of the day, and also how much was eaten.”.

3. What's your fMRI sequence parameters? In general, the TR of fMRI for 3T is 2s. If this, each image should be presented for 1 TR and with an ISI 1 TR at least.

Reply We agree with the reviewer's comment that for a TR of 2s (which will be the case in our measurement), image presentation should be at least 1 TR, which is true for an event-related design (for the interest of neuronal activation on single images). Our protocol intends to use a block-design with total net-stimulation duration of 13s per block.
To cope with the raised issues, we have therefore now done the following:
(1.) We added a statement that we will use a block-design: “We decided to use a block design fMRI experiment with 1 s of image presentation and a brief ISI to maximize the neural responses to the visual stimuli. Our stimulus duration of 1 s is shorter than the stimulus duration used e.g. by Blechert et al.42 who presented their images for 2.5 s in an event-related paradigm. The shorter stimulus duration of 1 s helps to avoid excessive eye movement.”

4. In the MRI data analysis, the timing correction should be before the motion correction. And the spatial normalization step should be after the head motion correction but before the spatial smooth. In line 15 to 18, what do you mean by saying " The temporal derivative of the hemodynamic response function will be included as regressor of no interest to model" ? I suggest to change it to "Some

nuisance covariates such as head motion parameters, global mean signal, white matter signal, and cerebrospinal fluid signal were regress out to reduce the no interest signals."

Reply: We agree with the reviewer's comment on the general guideline of fMRI-data analysis. Depending on the software used for data analyzing, different recommendations are available: slice timing correction is for instance recommended when using statistic parametric mapping (SPM), but is rarely used in analysis with FSL. We specify the statement in line 15 to 18 by including your suggestion.

To cope with the raised issues, we have therefore now done the following:
(1.) We slightly modified the sentence to:" For preprocessing, functional data will be motion corrected, temporally filtered with a high-pass filter, and spatially smoothed using a Gaussian kernel."
(2.) We changed the sentence (line 15-18) to: "Some nuisance covariates such as head motion parameters, global mean signal, white matter signal, and cerebrospinal fluid signal will be regressed out to reduce the signal of no interest."

5. In line 27 to 28, "For thresholding and correction for multiple comparisons, cluster-wise thresholding will be used" The thresholding is for multiple comparisons correciton. What's your thresholding? I suggest to change it to "For correction of multiple comparisons, clusters were threshold at (your value). And put this sentence back to the group comparisons. Also, please give out your P values.
6, Please noted the order of the data analysis steps.

Reply: We intend to threshold with a Z-value of 3.1 ($p<0.05$).

To cope with the raised issues, we have therefore now done the following:
(1.) We changed the sentence structure and added the values for thresholding:
" For correction of multiple comparisons, clusters will be thresholded at $Z = 3.1$ ($p<0.05$)."
Literature:
1. van der Laan, L. N., De Ridder, D. T., Viergever, M. A., & Smeets, P. A. (2011). The first taste is always with the eyes: a meta-analysis on the neural correlates of processing visual food cues. Neuroimage, 55(1), 296-303.
2. Morínigo, R., Moizé, V., Musri, M., Lacy, A. M., Navarro, S., Marín, J. L., ... & Vidal, J. (2006). Glucagon-like peptide-1, peptide YY, hunger, and satiety after gastric bypass surgery in morbidly obese subjects. The Journal of Clinical Endocrinology & Metabolism, 91(5), 1735-1740.

Reviewer#2
Dear Mr. Lyle Wiemerslage
Dear Mrs. Gaia Olivo

1. The protocol was unclear as to if the patients will be fasted or sated during scanning? This is an important factor to consider as past studies have found differences between prandial states. In any case, hunger should be assessed prior to scanning and prandial state should be carefully controlled between all groups. Moreover, fasting should not be equated with hunger. Just because someone fasted overnight, does not mean they are hungry. Perhaps scans in the early afternoon after at least a 4-hour fast are a superior fasting/hungry condition compared to overnight fasts. We shouldn't necessarily measure hunger in the brain the same way we measure blood sugar. I encourage the authors to consider what exactly are the phenomena of fasting versus hunger, and when each is best displayed/measurable: time of day, time from previous meal, size of previous meal, etc. Likewise, if the authors intend to scan in a sated state, then they carefully consider the meal size, timing, palatability, etc. Some studies have used small, set meal sizes, while other studies have employed meals with caloric contents based on body size. I again encourage the authors to likewise think

deeply about what “sated” means and how it is best displayed/measured for the purposes of their study.
Reply: We appreciate your comment on the differentiation between fasted and sated and its influence on food processing. Based on your concerns and on findings in a meta-analysis1, we decided to extend the fasted period to 4 hours and all measurements will take place at around the same time. We also realize that these stages are not well described in our manuscript
To cope with the raised issues, we have therefore now done the following:
(1.) We made statements about the fasting/hungry conditions more explicit by adding the paragraph “The participants will be in a fasted state 4 hours prior to measurement. All the measurements will take place at around the same time of the day, which will be between 3pm, and 5pm. The current state of hunger at the recording time will be assessed on a ten-point-Likert scale. Additionally, the time from the previous meal and portion size of previous meal, will be recorded. To have a measurement on food consumption before entering the study, the participants will be asked to complete a food diary 1 week prior to the study in order to record what has been eaten, at which time of the day, and also how much was eaten.”

2. Have the authors considered scanning in both fasting and sated states (both before and after surgery)? It is twice the scanning, but it exponentially increases the potential findings of the study. The resting-state fMRI would be the most important to consider for these extra scans.
Reply: We are mainly interested in the longitudinal effects of RYGB on food processing and executive functioning by answering the question of long-term-relationship between RYGB and food processing. By applying the same protocol on the pre-and post-measurement we believe it is sufficient to explain the influence of RYGB on neuronal food processing and executive functioning, indeed it would be an interesting study to explore the variables of fasting and sated states within this background.

3. Previous studies have found sometimes conflicting results between males and females in fMRI studies regarding food cues. And with only 25 obese participants planned, it seems prudent to consider enrolling only one sex. Admittedly, this is not ideal, but it simplifies the analysis and leaves less room for argument. I would consider females the more interesting/important choice of study (as males have been typically preferred in past studies). Regardless of what the authors decide, the female hormonal cycle should be controlled unless the authors can adequately explain that it is not expected to be an issue. In addition, if the authors ignore this complaint, they should be careful to split the balance between males and females.
Reply: We fully agree with the reviewers and decided to only study female participants. To control for the menstrual cycle, we decided to have the measurements up to 4 weeks prior and up to 4 weeks after the RYGB to control for the menstrual cycle.

To cope with the raised issues, we have therefore now done the following:
(1.) We explicitly stated that we will control for the menstrual cycle of the female participants: ” Since the study will only enroll female participants the menstrual cycle will be controlled for by only include females in the midfollicular phase (days 4-8 after onset of menses) due to differences in brain responses to reward between the follicular and the luteal phase”.

4. The inclusion/exclusion criteria are rather vague:
a. Could the authors provide the full list of inclusion/exclusion criteria?
b. Will medication status be considered a criterion? Psychopharmacological treatment has been mentioned as an exclusion criterion, but what about medications that might affect hormonal levels – a secondary outcome of this study?
c. Participants will be assessed for comorbidities by administering the SKID-I. But how will the results of this assessment be used? In case of comorbidities, will the participant be included in the study all the same, or will he/she be rejected?
d. Neurological disorders should be an exclusion criterion.

Reply:
a) The inclusion criteria are:
Age: 18-60
BMI: >35kg/m2 (Obesity Group); <25kg/m2 (Control Group)
Gender: Female
Participants in the obesity group will receive RYGB operation

The exclusion criteria are:
Substance abuse
Smoking
Psychiatric disorders (except anxiety in the obese-group)
History of eating disorder
Psychopharmacological treatment
Pregnancy
Claustrophobia
Neurological disorders

b) We see the problem with medication that will affect hormonal status and thank the reviewer for highlighting it. Nevertheless, we decided that this kind of medication will not be an exclusion criterion. It might be the case that participants take these medicaments at both the pre-and post-measurement, which would also allow us to capture a possible change (given a stable dose) in neuronal food processing. If the dose is not stable or the treatment is discontinued, we would consider this statistically (by applying a regression model in order to rule out the effect of medication, depending on the number of participants), or to exclude the participant post-hoc.
c) The SKID-I is mainly used for assessing affective disorders, psychotic disorders, somatoform disorders, anxiety disorders, adjustment disorders, and eating disorders (bulimia nervosa, anorexia nervosa). Except from anxiety disorders in the obese patients group, the presence of any other comorbid disorder will lead to exclusion of the study. For assessing personality disorders the SKID-II will also be done.
d) We agree at this point and added this to the exclusion criteria.

5. Haven't hypotheses #1 and 2 already been demonstrated in previous works? How exactly will this study build on these past findings?
Reply: Our main research question is, whether the changes associated with RYGB in neuronal food-stimuli processing (which indeed are well documented for shorter periods of time) are prolonged to the period of one year after RYGB. As far as we know, there is no study investigating such a possible prolonged change in neuronal food-processing in this specific period of time (an exception might be Wang et al. 2016 who studied the change of actual taste 12 month after RYGB, however they included only five participants (3 males))2. We believe that we can contribute to the existing literature by closing the gap of food-stimuli processing and working memory performance one year after RYGB. Because hypothesis #1 and 2 are inevitable linked (at first, we must investigate the presence of aberrant neuronal food stimuli processing in this specific population, in order to hypothesize an adjustment to the normal control group one year after RYGB), we believe hypothesis #1is legit (despite that it has been shown in past studies).

6. Regarding hypotheses #3 and 4, I don't understand why memory performance is of concern (other than that it has been tested previously). Eating behaviour is the primary psychological element we are interested in. So shouldn't tests specific to eating behaviour be considered primary endpoints?
Indeed, working memory performance and executive functions (as measured via the Trail Making Test; J Clin Exp Neuropsychol. 2000 Aug;22(4):518-28.) have both been found to be affected in obesity by meta-analyses (Cook, Rebecca et al. Obesity Research & Clinical Practice , Volume 8 , 21; Veronese et al., Neurosci Biobehav Rev. 2017 Jan;72:87-94. doi: 10.1016/j.neubiorev.2016.11.017.).

Reply: We absolutely agree with the reviewers' comment that eating behavior is the primary aim for investigation in the study. Furthermore, we want to investigate the effects of RYGB on working memory on a neuronal level with a 12 month follow-up period as this has not been done in the current literature. We realized that our hypothesis overview in Box 1 concerning the working memory performance must be extended to neuronal assumptions, otherwise one could have the impression that we investigate working memory performance on a behavioral level only (which indeed has been done before).

To cope with the raised issues, we have therefore now done the following:
(1.) In Box 1, we revised hypotheses #3 and 4 by adding our assumption regarding the neuronal outcomes of working memory associated with obesity and RYGB:" H3: Working memory performance will be lower in obese patients prior to RYGB compared to healthy controls, which is reflected in lower prefrontal activation.
H4: Working memory performance will be improved 12 months after RYGB in obese patients, which is reflected in increased prefrontal activation."

7. Hypothesis #6 is too vague. How exactly do the authors expect hormones to correlate with the other variables measured? Also, what exactly is of consequence in these correlations regarding behaviour or neural activity? How will these insights lead to improved treatments?
Reply: We reformulated hypothesis #6 by pointing out that we expect a negative relationship between levels of GLP-1 and PYY and the neuronal activation in reward-related brain areas during food-stimuli processing. More specifically, we expect that lower hormonal levels are negatively correlated to a heightened brain response in reward-related areas to food stimuli prior to RYGB. Respectively, we expect that higher hormonal levels are negatively correlated with a lowered brain response in reward areas to food stimuli 12 months after RYGB.
To cope with the raised issues, we have therefore now done the following:
(1.) We included a more detailed statement in the study aims-section:" The hormonal status of PYY and GLP-1 are expected to be negatively correlated with the neuronal response in reward related brain areas to food images. We hypothesize that lower hormonal level are negatively correlated to a heightened brain response in reward related areas to food stimuli prior to RYGB. Respectively, we expect higher hormonal levels are negatively correlated with a lowered brain response in reward areas to food stimuli 12 month after RYGB."
(2.) In Box 1 we specified H6:"Hormonal status is negatively correlated with the neuronal response in reward related brain areas to food images."

8. Why is the follow up 12 months? I suppose it controls for the time of year, but shouldn't results from the surgery be measurable before that? Some studies have seen differences in neural activity as early as one month after surgery (perhaps that's too soon after the recovery). But the goal is to simply measure patients after they have lost weight (or failed to lose weight), correct? Why needlessly make the study longer than it needs to be and potentially lose patients to follow up? Indeed, a year after surgery may be too long – perhaps some patients have an initial weight loss and then rebound. The literature should be consulted for an optimal follow-up time point. Perhaps weight can be tracked via a patient diary?

Reply: We are interested whether a neuronal change in food-stimuli processing and working memory functioning is present on a longitudinal basis. The current literature investigated changes after RYGB for different time-periods, including shorter periods (most often at a behavioral level only; to give the reader an overview of studies, we included table 1 in the manuscript), however it is not known, whether these changes are long-lasting up to one year after RYGB. Scholtz et al. 2013 could show a long-lasting change in hedonic brain responses to food ~8 months after RYGB3. A study measuring potential long-lasting effects 12 months after RYGB is lacking. You might be right in the case that

some patients can gain weight again after a certain time after the operation. In order to capture these fluctuations in weight, we see the necessity to include a patient diary in which the participants have to protocol their weight on a weekly basis. On the neuronal level, it might be intuitive that a patient who regains weight after a successful weight-reduction associated with RYGB shows the same neuronal activation patterns found prior to RYGB. On the other hand, it might also be the case that the brain responses are compared to normal, lean controls that would allow the conclusion of a brain-behavior dysbalance in the case of weight fluctuations after RYGB. We believe that it is worth to investigate the neuronal change with especially the time-interval of 12 months after RYGB and therefore add knowledge of the neuronal change induced by RYGB to the community.
To cope with the raised issues, we have therefore now done the following:
(1.) We included the sentence:" Also, to have a measurement on weight after RYGB, the participants need to record the body-weight on a weekly basis during the pre-post-interval."

9. The authors should consider collecting blood samples for future genetic analysis, as there are several other studies in the literature that they could compare results. Variants for the FTO gene have been the most popular to date.

Reply: We appreciate the idea. We considered collecting an additional sample of blood that will be stored for further genetic analysis. Presumably, we report the collecting of the additional blood sample in the respective result article.

10. Regarding the images shown to the participants during the fMRI task, how exactly are the images chosen? Are the food images highly palatable, equal in caloric content, macronutrients, taste, perceived healthiness, etc.? Likewise, are the control images matched for features relative to the food images: colors, shapes, etc.? How exactly is it determined that a food is high-calorie versus low-calorie? Also, something can be high-calorie but still perceived as relatively healthy, which may not be the intended stimulus to the participant. The authors should consider what feelings/associations will be intentionally or unintentionally elicited to the participants. For example, an image of a birthday cake may remind someone that their birthday was forgotten and unintentionally illicit a neural response based on sad emotions. Perhaps a 1-second interval is too fast for emotive processing, but that should be adequately explained, if so.

Reply: We did an in-house evaluation of 20 female co-workers and students who rated the pictures as either high- or low-caloric, as well as the arousal and valence with the self-assessment manikin (SAM)4. Concerning the 1-second picture interval: An event-related-potential study from Cuthbert et al. (2000)5 could show that emotional processing of pictures (selection from international affective picture system; IAPS) begins already 400-700ms after picture onset and lasts up to five seconds. Interestingly, the participants rated the figures with the SAM (similar to the pictures we will use) and a factor analysis on emotional valence and emotional arousal could account for 30.6% and 40% of the variance in the EEG dataset. Since we show 15 (1s each) pictures in a block design, we take advantage of the early emotional processing component on each picture, while also measuring the full hemodynamic response over the whole block.
To cope with the raised issues, we have therefore now done the following:
(1.) We explicitly wrote down that the pictures were rated according to the criteria given above," To account for possible differences on arousal and valence between pictures, we did an in-house rating: each picture was rated from 20 females with the self-assessment manikin (SAM) on the scales arousal and valence. Further, the food pictures were rated either as high- or low-caloric.

11. The independent-component analysis for the resting-state fMRI is sound, but is also a somewhat curious choice, considering there are so many brain areas that qualify for a priori testing. Perhaps that is why the authors chose this method (too many brain areas)? Regardless, I would be interested to know their reasoning/comparisons to the other available methods for analysing resting-state data.

Reply: We realized that our statement for analyzing the resting state might be too unspecific. We intend to do a seed-to-voxel analysis and independent component analysis (ICA). For the seed-to-voxel analysis method, a priori regions of interest will be chosen based on earlier findings (e.g., salience network). The ICA method does not use a priori assumptions and will be included in order to detect networks that have not been presumed before.
To cope with the raised issues, we have therefore now done the following:
(1.) We revised our statement on resting state fMRI:
" Resting state fMRI will be analyzed with the seed-to-voxel method using the toolbox CONN (http://www.conn-toolbox.org) and with independent component analysis using dual regression as implemented in FSL48.
For the seed-to-voxel analysis, regions of interest will be chosen a priori based on the findings of relevant earlier studies (e.g., salience network). Associations between the time courses of the fMRI signal in these seed regions and in all other voxels of the brain will be computed.
For dual regression, a model-free approach based on FSL's MELODIC will be used. All resting state data sets will be decomposed into sets of time courses and associated spatial maps which describe the temporal and spatial characteristics of underlying hidden signals47."

12. While the detailed description of the statistical analyses is appreciated, the authors should assess the validity of the assumptions before choosing the appropriate tests. This can only be done after data collection. For example, they mention that Student's t-test will be used for hormonal measurements. The t-test, as all parametric tests (ANOVA included) has several assumptions, mainly the normality of the residuals and homogeneity of variance between groups.

Reply: We agree that we cannot really make these assumptions without a look in the data set. We changed the sentence accordingly
To cope with the raised issues, we have therefore now done the following:
(1.) We added that we plan to calculate a repeated measures ANOVA given the circumstances of normal distribution: "If the data are normally distributed, we plan to do repeated measures analysis of variance (ANOVA) with the factors group X time will be used to determine group and time differences for the demographic, psychometric and neuropsychological data."
(2.) Differences in hormonal status will be presumably evaluated with Student's T-tests, given that the data are normally distributed.

13. Regarding the neuroimaging protocol, are the scans assessed by a neuroradiologist to ensure the absence of structural anomalies? This is usually a standard procedure, but still important to say so – especially for a "methods" manuscript such as this.
Reply: On each measurement, a MR technologist is present. We include this information accordingly in the manuscript.
To cope with the raised issues, we have therefore now done the following:
(1.) We added, "To ensure the absence of structural abnormalities in the brain, a medical technical radiological assistant will always be present during the measurements. In case of an abnormality, we will contact a radiologist for further diagnostic scanning."

14. Concerning the resting-state fMRI, the authors might consider having the participants fixating on a cross during the acquisition, as this protocol has been demonstrated to have a better test-retest reliability (Patriat, Neuroimage. 2013 Sep;78:463-73. doi: 10.1016/j.neuroimage.2013.04.013.). This could be useful, as this is a longitudinal study.
Reply: We thank the author for this useful information. We will include a fixation cross during the resting-state fMRI.
To cope with the raised issues, we have therefore now done the following:
(1.) We added," Resting-state fMRI will be measured for 9 min. The participants will be instructed to

not focus on specific thoughts and remain still with eyes open. The participants will be instructed to fixate a cross during the recording."

Literature

1. van der Laan, L. N., De Ridder, D. T., Viergever, M. A., & Smeets, P. A. (2011). The first taste is always with the eyes: a meta-analysis on the neural correlates of processing visual food cues. Neuroimage, 55(1), 296-303.
2. Wang J-L, Yang Q, Hajnal A, et al. A pilot functional MRI study in Roux-en-Y gastric bypass patients to study alteration in taste functions after surgery. Surgical endoscopy 2016;30(3):892-98.
3. Scholtz S, Miras AD, Chhina N, et al. Obese patients after gastric bypass surgery have lower brain-hedonic responses to food than after gastric banding. Gut 2013:gutjnl-2013-305008.
4. Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. Journal of behavior therapy and experimental psychiatry, 25(1), 49-59.
5. Cuthbert, B. N., Schupp, H. T., Bradley, M. M., Birbaumer, N., & Lang, P. J. (2000). Brain potentials in affective picture processing: covariation with autonomic arousal and affective report. Biological psychology, 52(2), 95-111.

VERSION 2 – REVIEW

REVIEWER	Baoci Shan Institute of High Energy Physics, China
REVIEW RETURNED	17-Jul-2018

GENERAL COMMENTS	In the manuscript " Impact of bariatric surgery on neural food processing and cognition - an fMRI study", the authors planed to investigate the relationship between neural activities and cognition as well as the impact of RYGB on these variables in the obese. It is an interesting and timely study. The protocol is detailed and feasible. Before publication, I have some minor questions: 1, In line 26, page 2, miss a full stop (e.g."connectivity Baseline...."). 2, Please give the references for findings in previous fMRI studies (from line 47 to 55, in page 4). 3, What do you mean by saying " Respectively, we expect higher hormonal levels are negatively corrected with a lowered brain response in reward areas to food stimuli 12 month after RYGB" ? 4, In line 37, page 11, please revised "will be also be obtained" to "will be also obtained". 5, Your MRI data analysis. The spatial normalization step is before the band pass filtering, smoothing, and regressing out nuisance covariates. Please revise this part carefully. 6, What do you mean by saying " we plan to perform repeated measures analyses of variance (ANOVA) with the factors group X time will be used to determine group and time differences for the demographic, psychometric and neuropsychological"? What is "X time"? Is a pronoun missed in this sentence?